

VT-d Posted Interrupts

Feng Wu, Jun Nakajima <Speaker> Intel Corporation

Agenda

- Motivation
- Difference btw CPU-based and VT-d Posted Interrupts
- Architecture
- Implementation Details
- Performance
- Summary



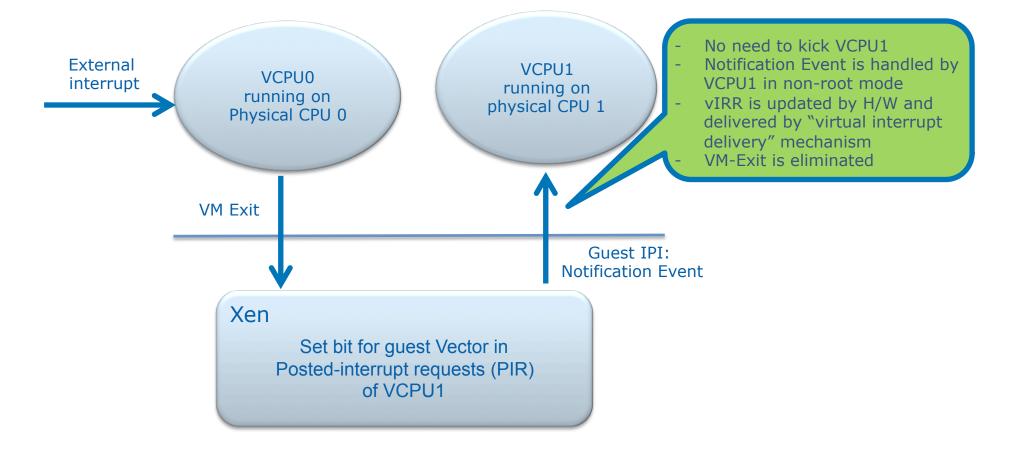
Motivation

- Interrupt virtualization efficiency
- Interrupt migration complexity
- Big requirement of host vector for different assigned devices



CPU-based Posted-Interrupt in Xen

• External interrupt handling





Key Data Structures for CPU-based Posted-Interrupt Processing

Posted-interrupt notification vector:

Send virtual interrupts to guests w/o VM exit
If physical vector == Posted-interrupt

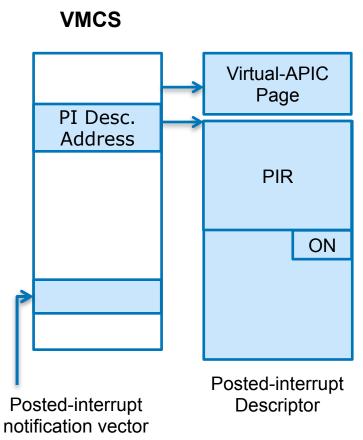
notification vector (VMCS field)

PIR (Posted-interrupt requests)

- Set bits for guest vectors in advance

ON (Outstanding Notification)

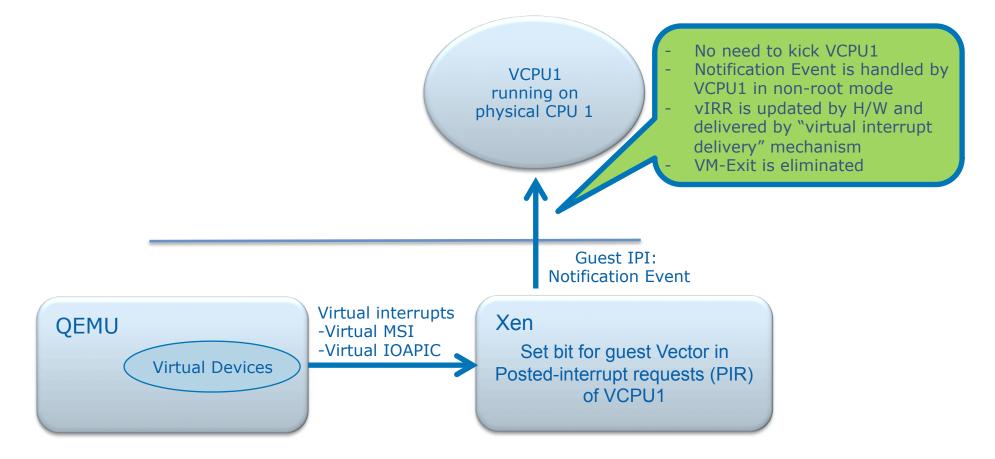
 If this bit is set, there is a notification outstanding for one or more posted interrupts





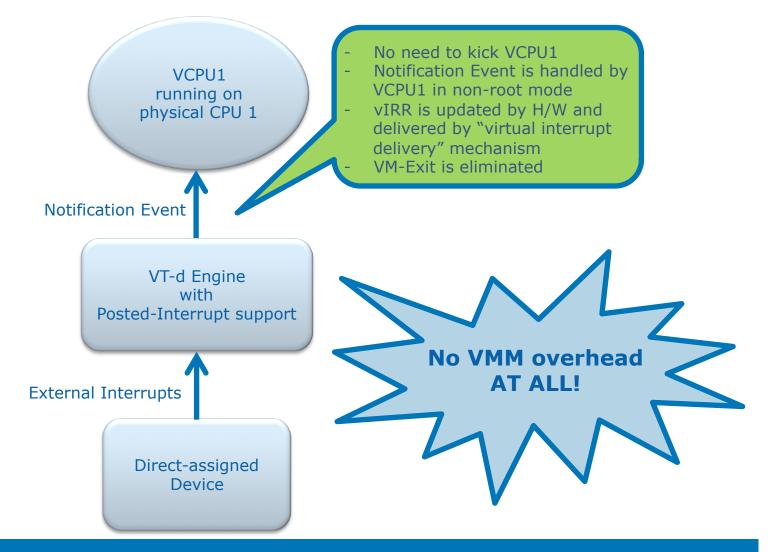
CPU-based Posted-Interrupt in Xen – cont'd

• Virtual interrupts from QEMU



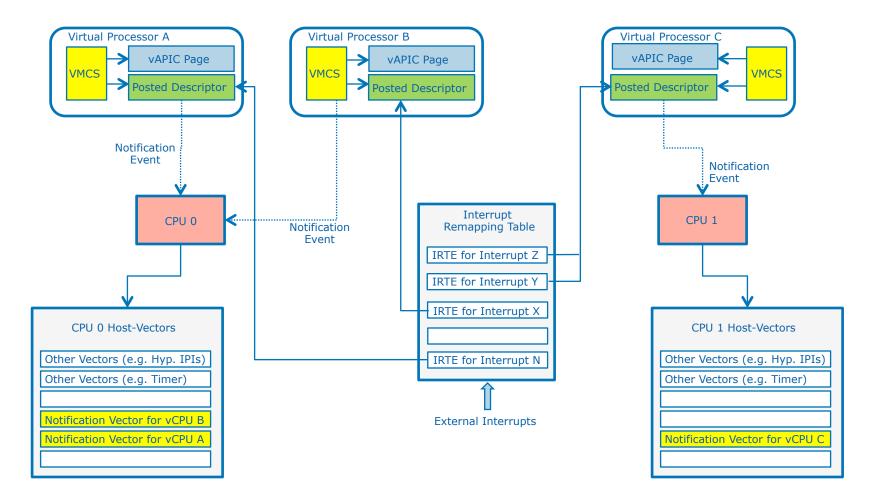


What's new for VT-d Posted-Interrupts





VT-d Posted-Interrupts Architecture



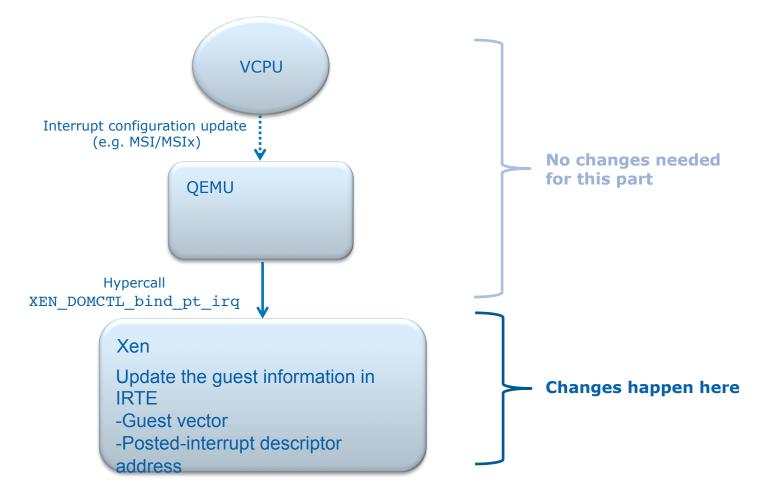


Xen Implementation Details:

- Update IRET according to guest's modification to the interrupt configuration (MSI address, data)
- Interrupt migration during VCPU scheduling



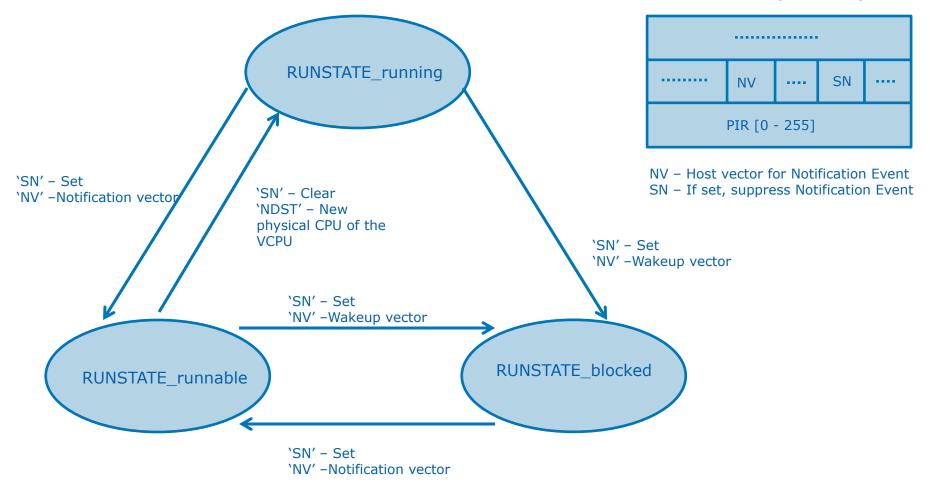
Xen Implementation: IRTE update





Xen Implementation: VCPU Scheduling

Posted-Interrupt Descriptor





Summary

- VT-d Posted-interrupts advantages
 - External interrupts from direct-assigned devices are delivered to guest running in non-root mode directly
 - Improve Interrupt virtualization efficiency, e.g. Less VM-Exits.
 - Simplify interrupt migration
 - Consume less physical interrupts
- Performance
- The Specification will be published very soon
 - Can be found in Intel website



Thank YOU! Q & A Or contact Feng Wu <feng.wu@intel.com>



Back up



VT-d Posted-Interrupts Support

- Interrupt-remap-table-entry (IRTE) enhanced as follows:
 - An existing reserved bit claimed to indicated Posted-interrupt (PST)
 - Software may choose to "remap" or "post" each interrupt independently
- IRTEs with 'PST' set are interpreted per below format

| 127 | Descriptor Address [63:32] | RSVD | S S V T T | | | | Sour | ce ID | | | 64 |
|-----|----------------------------|------|-------------------|-------------|----------|-------------|-------|-------|-------------|---|----|
| 63 | Descriptor Address [31:6] | RSVD | Virtual Vector | P S T | RS VD | U R G | AVAIL | RSVD | F P D | P | 00 |

- New Fields
 - Descriptor Address: the address of the posted-interrupt descriptor
 - Virtual Vector: the guest vector of the interrupt
 - URG: indicates if the interrupt is urgent
- Other fields continue to have the same meaning



Posted-interrupt Descriptor

64-byte aligned structure as follows

| 3Fh | RSVD | | | | | | 30h | | |
|-----|---------------|------|------|----|-----|------|-----|-----|-----|
| 2Fh | RSVD | NDST | RSVD | NV | NDM | RSVD | SN | ÔN | 20h |
| 1Fh | PIR [255:128] | | | | | | | 10h | |
| 0Fh | PIR [127:0] | | | | | | | 00h | |

Fields

| Posted Interrupt Requested (PIR) | Storage to "post" interrupts to a VP (1-bit per virtual-vector) | | | | |
|-------------------------------------|---|--|--|--|--|
| Outstanding Notification (ON) | If set, indicates pending notification event yet to be serviced (no need to send another) | | | | |
| Suppress Notification (SN) | If set, suppress notification event when posting non-urgent interrupts | | | | |
| Notification Destination Mode (NDM) | Indicates Logical/Physical destination-mode for notification event; Delivery mode and level forced to Fixed/Edge | | | | |
| Notification Vector (NV) | Host-vector for the notification event | | | | |
| Notification Destination (NDST) | APIC-ID of target CPU for notification event (8-bits in xAPIC mode; 32-bits in x2APIC mode) | | | | |



VT-d: Steps for Interrupt Posting

